

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: HOFVANDER, Per
PERSSON, Per T
WIKSTROM, Olle
TALLBERG, Anneli

(ii) TITLE OF INVENTION: GENETICALLY ENGINEERED MODIFICATION OF
POTATO TO FORM AMYLOPECTIN-TYPE STARCH

(iii) NUMBER OF SEQUENCES: 21

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Burns, Doane, Swecker & Mathis
(B) STREET: George Mason Bldg., Washington & Prince Sts.
(C) CITY: Alexandria
(D) STATE: Virginia
(E) COUNTRY: United States
(F) ZIP: 22313-1404

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/070,455
(B) FILING DATE: 09-JUN-1993
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Crane-Feury, Sharon E
(B) REGISTRATION NUMBER: 36,113
(C) REFERENCE/DOCKET NUMBER: 003300-293

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (703) 836-6620
(B) TELEFAX: (703) 836-2021

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 217..342

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGCATGTTTC CCTACATTCT ATTTAGAACATC GTGTTGTGGT GTATAAACGT TGTTTCATAT 60
CTCATCTCAT CTATTCTGAT TTTGATTCTC TTGCCTACTG TAATCGGTGA TAAATGTGAA 120
TGCTTCCTTT CTTCTCAGAA ATCAATTCT GTTTGTTTT TGTTCATCTG TAGCTTATTG 180

RECEIVED
JULY 02 1993
GROUP 1800

TCTGGTAGAT TCCCCTTTT GTAGACCACA CATCAC ATG GCA AGC ATC ACA GCT Met Ala Ser Ile Thr Ala	234
1 5	
TCA CAC CAC TTT GTG TCA AGA AGC CAA ACT TCA CTA GAC ACC AAA TCA Ser His His Phe Val Ser Arg Ser Gln Thr Ser Leu Asp Thr Lys Ser	282
10 15 20	
ACC TTG TCA CAG ATA GGA CTC AGG AAC CAT ACT CTG ACT CAC AAT GGT Thr Leu Ser Gln Ile Gly Leu Arg Asn His Thr Leu Thr His Asn Gly	330
25 30 35	
TTA AGG GCT GTT Leu Arg Ala Val	342
40	

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2549 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

B1
cont

AACAAGCTTG ATGGGCTCCA ATCAACAACT AATACTAAGG TAACACCCAA GATGGCATCC	60
AGAACTGAGA CCAAGAGACC TGGATGCTCA GCTACCATTG TTTGTGGAAA GGAAATGAAC	120
TTGATCTTG TGGGTACTGA GGTTGGCTCT TGGAGCAAAA CTGGTGGACT AGGTGATGTT	180
CTTGGTGGAC TACCACCAAGC CCTTGCAGTA AGTCTTCTT TCATTTGGTT ACCTACTCAT	240
TCATTACTTA TTTTGTCTAG TTAGTTCTA CTGCATCAGT CTTTTATCA TTTAGGCCCG	300
CGGACATCGG GTAATGACAA TATCCCCCG TTATGACCAA TACAAAGATG CTTGGGATAC	360
TGGCGTTGCG GTTGAGGTAC ATCTTCTAT ATTGATACGG TACAATATTG TTCTCTTACA	420
TTTCCTGATT CAAGAACATGTG ATCATCTGCA GGTCAAAGTT GGAGACAGCA TTGAAATTGT	480
TCGTTCTTT CACTGCTATA AACGTGGGT TGATCGTGT TTTGTTGACC ACCCAATGTT	540
CTTGGAGAAA GTAAGCATAT TATGATTATG AATCCGTCT GAGGGATACG CAGAACAGGT	600
CATTTGAGT ATCTTTAAC TCTACTGGTG CTTTTACTCT TTTAAGGTTT GGGGCAAAAC	660
TGGTTCAAAA ATCTATGGCC CCAAAGCTGG ACTAGATTAT CTGGACAATG AACTTAGGTT	720
CAGCTTGTG TGTCAAGTAA GTTAGTTACT CTTGATTTT ATGTGGCATT TTACTCTTT	780
GTCTTTAAC GTTTTTTAA CCTTGTTTC TCAGGCAGCC CTAGAGGCAC CTAAAGTTT	840
GAATTTGAAC AGTAGCAACT ACTTCTCAGG ACCATATGGT AATTAACACA TCCTAGTTTC	900
AGAAAAACTCC TTACTATATC ATTGTAGGTA ATCATCTTA TTTGCCTAT CCCTGCAGGA	960
GAGGATGTTG TCTTCATTGC CAATGATTGG CACACAGCTC TCATTCCTG CTACTGAAAG	1020
TCAATGTACC AGTCCAGAGG AATCTACTTG AATGCCAAGG TAAAATTCT TTGTATTAC	1080
TCGATTGCAC GTTACCCCTGC AAATCAGTAA GGTTGTATTA ATATATGATA AATTCACAT	1140

B1
cont

TGCCTCCAGG TTGCTTCCTG CATCCATAAC ATTGCCTACC AAGGTCGATT TTCTTTCTCT	1200
GACTTCCCTC TTCTCAATCT TCCTGATGAA TTCAGGGGTT CTTTGATTT CATTGATGGG	1260
TATGTATTTA TGCTTGAAAT CAGACCTCCA ACTTTGAAG CTCTTTGAT GCTAGTAART	1320
TGAGTTTTA AAATTTGCA GATATGAGAA GCCTGTTAAG GGTAGGAAAA TCAACTGGAT	1380
GAACGGCTGGG ATATTAGAAT CACATAGGGT GGTTACAGTG AGCCCATACT ATGCCCAAGA	1440
ACTTGTCTCT GCTGTTGACA AGGGAGTTGA ATTGGACAGT GTCCTCGTA AGACTTGCAT	1500
AACTGGGATT GTGAATGGCA TGGATACACA AGAGTGGAAC CCAGCGACTG ACAAAATACAC	1560
AGATGTCAAA TACGATATAA CCACTGTAAG ATAAGATTTT TCCGACTCCA GTATATACTA	1620
AATTATTTG TATGTTTATG AAATTAAGA GTTCTTGCTA ATCAAATCT CTATACAGGT	1680
CATGGACGCA AAACCTTAC TAAAGGAGGC TCTTCAAGCA GCAGTTGGCT TGCCTGTTGA	1740
CAAGAAGATC CCTTGATTG GCTTCATCGG CAGACTTGAG GAGCAGAAAG GTTCAGATAT	1800
TCTTGTTGCT GCAATTCAACA AGTCATCGG ATTGGATGTT CAAATTGTAG TCCTTGTAAG	1860
TACCAAATGG ACTCATGGTA TCTCTCTGT TGAGTTTACT TGTGCCAAA CTGAAATTGA	1920
CCTGCTACTC ATCCTATGCA TCAGGAACT GGCAAAAAGG AGTTGAGCA GGAGATTGAA	1980
CAGCTCGAAG TGTTGTACCC TAACAAAGCT AAAGGAGTGG CAAAATTCAA TGTCCCTTG	2040
GCTCACATGA TCACTGCTGG TGCTGATTTT ATGTTGGTTC CAAGCAGATT TGAACCTTGT	2100
GGTCTCATTG AGTTACATGC TATGCGATAT GGAACAGTAA GAACCAGAAG AGCTTGTACC	2160
TTTTACTGA GTTTTAAAAA AAAGAACAT AAGACCTTGT TTTCCATCTA AAGTTTAATA	2220
ACCAACTAAA TGTTACTGCA GCAAGCTTT CATTCTGAA AATTGGTTAT CTGATTTAA	2280
CGTAATCACA TGTGAGTCAG GTACCAATCT GTGCATCGAC TGGTGGACTT GTTGACACTG	2340
TGAAAAGAAGG CTATACTGGA TTCCATATGG GAGCCTTCAA TGTTGAAGTA TGTGATTTA	2400
CATCAATTGT GTACTTGTAC ATGGTCCATT CTCGTCTTGA TATACCCCTT GTTGCATAAA	2460
CATTAACCTTA TTGCTTCTTG AATTTGGTTA GTGCGATGTT GTTGACCCAG CTGATGTGCT	2520
TAAGATAGTA ACAACAGTTG CTAGAGCTC	2549

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 492 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..15

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 101..218

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAG CTC TCC TGG AAG GTAAGTGTGA ATTTGATAAT TTGCGTAGGT ACTTCAGTT 55
Glu Leu Ser Trp Lys
1 5

GTTGTTCTCG TCAGCACTGA TGGATTCCAA CTGGTGTCT TGCAG GAA CCT GCC 109
Glu Pro Ala
1

AAG AAA TGG GAG ACA TTG CTA TTG GGC TTA GGA GCT TCT GGC AGT GAA 157
Lys Lys Trp Glu Thr Leu Leu Leu Gly Leu Gly Ala Ser Gly Ser Glu
5 10 15

CCC GGT GTT GAA GGG GAA GAA ATC GCT CCA CTT GCC AAG GAA AAT GTA 205
Pro Gly Val Glu Gly Glu Ile Ala Pro Leu Ala Lys Glu Asn Val
20 25 30 35

GCC ACT CCT TAAATGAGCT TTGGTTATCC TTGTTTCAAC AATAAGATCA 254
Ala Thr Pro *

TTAAGCAAAC GTATTTACTA GCGAACTATG TAGAACCTA TTATGGGTC TCAATCATCT 314
ACAAAATGAT TGGTTTTGC TGGGGAGCAG CAGCATATAA GGCTGTAAAA TCCTGGTTAA 374
TGTTTTGTA GGTAAGGGCT ATTTAAGGTG GTGTGGATCA AAGTCAATAG AAAATAGTTA 434
TTACTAACGT TTGCAACTAA ATACTTAGTA ATGTAGCATA AATAACTA GAACTAGT 492

B1
cont X
(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 987 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAGCTTAAAC GAGATAGAAA ATTATGTTAC TCCGTTTGT TCATTACTTA ACAATGCAA 60
CAGTATCTTG TACCAAATCC TTTCTCTTT TTCAAACTTT TCTATTGGC TGTTGACGGA 120
GTAATCAGGA TACAAACCAC AAGTATTAA TTGACTCCTC CGCCAGATAT TATGATTTAT 180
GAATCCTCGA AAAGCCTATC CATTAAGTCC TCATCTATGG ATATACTTGA CAGTATCTTC 240
CTGTTGGGT ATTTTTTTT CCTGCCAAGT GGAACGGAGA CATGTTATGA TGTATACGGG 300
AAGCTCGTTA AAAAAAAATA CAATAGGAAG AAATGTAACA AACATTGAAT GTTGTTTTA 360
ACCATCCTTC CTTTAGCAGT GTATCAATT TGTAATAGAA CCATGCATCT CAATCTTAAT 420
ACTAAAATGC AACTTAATAT AGGCTAAACC AAGATAAAGT AATGTATTCA ACCTTTAGAA 480
TTGTGCATTC ATAATTAGAT CTTGTTGTC GTAAAAAATT AGAAAATATA TTTACAGTAA 540
TTTGGAAATAC AAAGCTAAGG GGGAAAGTAAC TAATATTCTA GTGGAGGGAG GGACCAGTAC 600
CAGTACCTAG ATATTATTT TAATTACTAT AATAATAATT TAATTAACAC GAGACATAGG 660
AATGTCAAGT GGTAGCGTAG GAGGGAGTTG GTTGTAGTTT TTAGATACTA GGAGACAGAA 720

240

CCGGACGGCC CATTGCAAGG CCAAGTTGAA GTCCAGCCGT GAATCAACAA AGAGAGGGCC	780
CATAATACTG TCGATGAGCA TTTCCCTATA ATACAGTGTC CACAGTTGCC TTCTGCTAAG	840
GGATAGCCAC CCGCTATTCT CTTGACACGT GTCACTGAAA CCTGCTACAA ATAAGGCAGG	900
CACCTCCTCA TTCTCACTCA CTCACTCACA CAGCTCAACA AGTGGTAACT TTTACTCATC	960
TCCTCCAATT ATTTCTGATT TCATGCA	987

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4964 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

B71
Unt

AAGCTTTAAC GAGATAGAAA ATTATGTTAC TCCGTTTGT TCATTACTTA ACAATGCAA	60
CAGTATCTTG TACCAAATCC TTTCTCTCTT TTCAAACCTT TCTATTGGC TGTTGACGGA	120
GTAATCAGGA TACAAACCAC AAGTATTAA TTGACTCCTC CGCCAGATAT TATGATTTAT	180
GAATCCTCGA AAAGCCTATC CATTAAGTCC TCATCTATGG ATATACTTGA CAGTATCTTC	240
CTGTTGGGT ATTTTTTTT CCTGCCAAGT GGAACGGAGA CATGTTATGA TGTATACGGG	300
AAGCTCGTTA AAAAAAAATA CAATAGGAAG AAATGTAACA AACATTGAAT GTTGTTTTA	360
ACCATCCTTC CTTTAGCAGT GTATCAATT TGTAATAGAA CCATGCATCT CAATCTTAAT	420
ACTAAAATGC AACTTAATAT AGGCTAAACC AAGATAAAGT AATGTATTCA ACCTTTAGAA	480
TTGTGCATTC ATAATTAGAT CTTGTTGTC GTAAAAAATT AGAAAATATA TTTACAGTAA	540
TTTCCAATAC AAAGCTAAGG GGGAAAGTAAC TAATATTCTA GTGGAGGGAG GGACCAGTAC	600
CAGTACCTAG ATATTATTT TAATTACTAT AATAATAATT TAATTAACAC GAGACATAGG	660
AATGTCAAGT GGTAGCGTAG GAGGGAGTTG GTTAGTTTT TTAGATACTA GGAGACAGAA	720
CCGGACGGCC CATTGCAAGG CCAAGTTGAA GTCCAGCCGT GAATCAACAA AGAGAGGGCC	780
CATAATACTG TCGATGAGCA TTTCCCTATA ATACAGTGTC CACAGTTGCC TTCTGCTAAG	840
GGATAGCCAC CCGCTATTCT CTTGACACGT GTCACTGAAA CCTGCTACAA ATAAGGCAGG	900
CACCTCCTCA TTCTCACTCA CTCACTCACA CAGCTCAACA AGTGGTAACT TTTACTCATC	960
TCCTCCAATT ATTTCTGATT TCATGCATGT TTCCCTACAT TCTATTATGA ATCGTGTGTT	1020
GGTGTATAAA CGTTGTTCA TATCTCATCT CATCTATTCT GATTTGATT CTCTTGCCTA	1080
CTGTAATCGG TGATAAATGT GAATGCTTCC TTTCTTCTCA GAAATCAATT TCTGTTTGT	1140
TTTGTTCAT CTGTAGCTTA TTCTCTGGTA GATTCCCCCTT TTTGTAGACC ACACATCACA	1200
TGGCAAGCAT CACAGCTTCA CACCACTTG TGTCAGAAG CCAAACCTCA CTAGACACCA	1260
AATCAACCTT GTCACAGATA GGACTCAGGA ACCATACTCT GACTCACAAT GGTTAAGGG	1320

CTGTTAACAA GCTTGATGGG CTCCAATCAA CAACTAATAC TAAGGTAACA CCCAAGATGG 1380
CATCCAGAAC TGAGACCAAG AGACCTGGAT GCTCAGCTAC CATTGTTGT GGAAAGGGAA 1440
TGAACTTGAT CTTTGTGGGT ACTGAGGTTG GTCCTTGGAG CAAAACGTGGT GGACTAGGTG 1500
ATGTTCTTGG TGGACTACCA CCAGCCCTTG CAGTAAGTCT TTCTTCATT TGGTTACCTA 1560
CTCATTCAATT ACCTTATTTG TTCTAGTTAGT TTCTACTGCA TCAGTCTTT TATCATTAG 1620
GCCCGCGGAC AGCGGGTAAT GACAATATCC CCCCGTTATG ACCAATACAA AGATGCTTGG 1680
GATACTGGCG TTGCGGTTGA GGTACATCTT CCTATATTGA TACGGTACAA TATTGTTCTC 1740
TTACATTTCC TGATTCAAGA ATGTGATCAT CTGCAGGTCA AAGTTGGAGA CAGCATTGAA 1800
ATTGTTCGTT TCTTCACTG CTATAAACGT GGGGTTGATC GTGTTTTGT TGACCACCCA 1860
ATGTTCTTGG AGAAAGTAAG CATATTATGA TTATGAATCC GTCCTGAGGG ATACGCAGAA 1920
CAGGTCAATT TGAGTATCTT TTAACTCTAC TGGTGCTTT ACTCTTTAA GGTTTGGGC 1980
AAAAACTGGTT CAAAAATCTA TGGCCCCAAA GCTGGACTAG ATTATCTGGA CAATGAACCTT 2040
AGGTTCAGCT TGTTGTGTCA AGTAAGTTAG TTACTCTTGA TTTTTATGTG GCATTTTACT 2100
CTTTTGTCTT TAATCGTTT TTTAACCTTG TTTTCTCAGG CAGCCCTAGA GGCACCTAAA 2160
GTTTGAAATT TGAACAGTAG CAACTACTTC TCAGGACCAT ATGGTAATTA ACACATCCTA 2220
GTTTCAGAAA ACTCCTTACT ATATCATTGT AGGTAATCAT CTTTATTTG CCTATTCCCTG 2280
CAGGAGAGGA TGTTCTCTTC ATTGCCAATG ATTGGCACAC AGCTCTCATT CCTTGCTACT 2340
TGAAGTCAAT GTACCAAGTCC AGAGGAATCT ACTTGAATGC CAAGGTAAAA TTCTTGTGTA 2400
TTCACTCGAT TGACACGTTAC CCTGCAAATC AGTAAGGTTG TATTAATATA TGATAAAATT 2460
CACATTGCCT CCAGGTTGCT TTCTGCATCC ATAACATTGC CTACCAAGGT CGATTTCTT 2520
TCTCTGACTT CCCTCTTCTC AATCTTCTG ATGAATTCAAG GGGTTCTTT GATTCATTG 2580
ATGGGTATGT ATTTATGCTT GAAATCAGAC CTCCAACCTT TGAAGCTCTT TTGATGCTAG 2640
TAAATTGAGT TTTTAAATT TTGCAGATAT GAGAAGCCTG TTAAGGGTAG GAAAATCAAC 2700
TGGATGAAGG CTGGGATATT AGAACATCACAT AGGGTGGTTA CAGTGAGCCC ATACTATGCC 2760
CAAGAACCTG TCTCTGCTGT TGACAAAGGAA GTTGAATTGG ACAGTGTCCCT TCGTAAGACT 2820
TGCATAACTG GGATTGTGAA TGGCATGGAT ACACAAGAGT GGAACCCAGC GACTGACAAA 2880
TACACAGATG TCAAATACGA TATAACCACT GTAAGATAAG ATTTTCCGA CTCCAGTATA 2940
TACTAAATTAA TTTTGTATGT TTATGAAATT AAAGAGTTCT TGCTAATCAA AATCTCTATA 3000
CAGGTCATGG ACGCAAAACC TTTACTAAAG GAGGCTCTTC AAGCAGCAGT TGGCTTGCCT 3060
GTTGACAAGA AGATCCCTT GATTGGCTTC ATCGGCAGAC TTGAGGAGCA GAAAGGTTCA 3120
GATATTCTTG TTGCTGCAAT TCACAAGTTC ATCGGATTGG ATGTTCAAAT TGTAGTCCTT 3180
GTAAGTACCA AATGGACTCA TGGTATCTCT CTTGTTGAGT TTACTTGTGC CGAAACTGAA 3240
ATTGACCTGC TACTCATCCT ATGCATCAGG GAACTGGCAA AAAGGATTTT GAGCAGGAGA 3300
TTGAACAGCT CGAAGTGTG TACCCTAACAA AAGCTAAAGG AGTGGCAAAA TTCAATGTCC 3360

342

CTTTGGCTCA CATGATCACT GCTGGTGCTG ATTTTATGTT GGTTCCAAGC AGATTTGAAC 3420
CTTGTGGTCT CATTCAAGTTA CATGCTATGC GATATGGAAC AGTAAGAACC AGAAGAGCTT 3480
GTACCTTTT ACTGAGTTT TAAAAAAAGA ATCATAAGAC CTTGTTTCC ATCTAAAGTT 3540
TAATAACCAA CTAAATGTTA CTGCAGCAAG CTTTCATTT CTGAAAATTG GTTATCTGAT 3600
TTAACGTAA TCACATGTGA GTCAGGTACC AATCTGTGCA TCGACTGGTG GACTTGTGA 3660
CACTGTGAAA GAAGGCTATA CTGGATTCCA TATGGGAGCC TTCAATGTTG AAGTATGTGA 3720
TTTTACATCA ATTGTGTACT TGTACATGGT CCATTCTCGT CTTGATATAC CCCTTGTGC 3780
ATAAACATTA ACTTATTGCT TCTTGAATTT GGTTAGTGCG ATGTTGTGA CCCAGCTGAT 3840
GTGCTTAAGA TAGTAACAAC AGTTGCTAGA GCTCTTCAG TCTATGGCAC CCTCGCATT 3900
GCTGAGATGA TAAAAAAATTG CATGTCAGAG GAGCTCTCCT GGAAGGTAAG TGTGAATTG 3960
ATAATTGCG TAGGTACTTC AGTTTGTGT TCTCGTCAGC ACTGATGGAT TCCAACGGT 4020
GTTCTTGCAG GAACCTGCCA AGAAATGGGA GACATTGCTA TTGGGCTTAG GAGCTTCTGG 4080
CAGTGAACCC GGTGTTGAAG GGGAAAGAAAT CGCTCCACTT GCCAAGGAAA ATGTAGGCCAC 4140
TCCTTAAATG AGCTTGGTT ATCCTTGTTC CAACAATAAG ATCATTAAAGC AAACGTATTT 4200
ACTAGCGAAC TATGTAGAAC CCTATTATGG GGTCTCAATC ATCTACAAAA TGATTGGTTT 4260
TTGCTGGGGA GCAGCAGCAT ATAAGGCTGT AAAATCCTGG TTAATGTTT TGTAGGTAAG 4320
GGCTATTTAA GGTGGTGTGG ATCAAAGTCA ATAGAAAATA GTTATTACTA ACGTTGCAA 4380
CTAAATACTT AGTAATGTAG CATAAATAAT ACTAGAACTA GTAGCTAATA TATATGCGTG 4440
AATTTGTTGT ACCTTTCTT GCATAATTAT TTGCACTACA TATATAATGA AAATTACCCA 4500
AGGAATCAAT GTTTCTTGCT CCGTCCTCCT TTGATGATTT TTTACGCAAT ACAGAGCTAG 4560
TGTGTTATGT TATAAATTGTT GTTAAAAGA AGTAATCAAA TTCAAATTAG TTGTTGGTC 4620
ATATGAAAGA AGCTGCCAGG CTAACTTGA GGAGATGGCT ATTGAATTTC AAAATGATTA 4680
TGTAAAACA ATGCAACATC TATGTCAATC AACACTTAAA TTATTGCATT TAGAAAGATA 4740
TTTTGAGCC CATGACACAT TCATTCAAA AGTAAGGTAG TATGTATGAT TGAATGGACT 4800
ACAGCTCAAT CAAAGCATCT CCTTACATA ACGGCACGT GTCTTGTCTA CTACTCTATT 4860
GGTAGTAGTA GTAGTAATTT TACAATCCAA ATTGAATAGT AATAAGATGC TCTCTATTTA 4920
CTAAAGTAGT AGTATTATTC TTTCGTTACT CTAAAGCAAC AAAA 4964

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (ix) FEATURE:
(A) NAME/KEY: Modified-site

- (B) LOCATION: 1..69
(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1-207 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Asn Lys Leu Asp Gly Leu Gln Ser Thr Thr Asn Thr Lys Val Thr Pro
1 5 10 15
Lys Met Ala Ser Arg Thr Glu Thr Lys Arg Pro Gly Cys Ser Ala Thr
20 25 30
Ile Val Cys Gly Lys Gly Met Asn Leu Ile Phe Val Gly Thr Glu Val
35 40 45
Gly Pro Trp Ser Lys Thr Gly Gly Leu Gly Asp Val Leu Gly Gly Leu
50 55 60
Pro Pro Ala Leu Ala
65

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 1..27
(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 296-377 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Ala Arg Gly His Arg Val Met Thr Ile Ser Pro Arg Tyr Asp Gln Tyr
1 5 10 15
Lys Asp Ala Trp Asp Thr Gly Val Ala Val Glu
20 25

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 1..33
(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 452-550 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Val Lys Val Gly Asp Ser I1 Glu Ile Val Arg Phe Phe His Cys Tyr
1 5 10 15

Lys Arg Gly Val Asp Arg Val Phe Val Asp His Pro Met Phe Leu Glu
20 25 30

Lys

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..30
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 647-736 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Val Trp Gly Lys Thr Gly Ser Lys Ile Tyr Gly Pro Lys Ala Gly Leu
1 5 10 15

Asp Tyr Leu Asp Asn Glu Leu Arg Phe Ser Leu Leu Cys Gln
20 25 30

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..21
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 815-878 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Ala Ala Leu Glu Ala Pro Lys Val Leu Asn Leu Asn Ser Ser Asn Tyr
1 5 10 15

Phe Ser Gly Pro Tyr
20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amin acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 1..34
 - (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 878 and 959-1059 of SEQ ID NO. 2."
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Gly Glu Asp Val Leu Phe Ile Ala Asn Asp Trp His Thr Ala Leu Ile
1 5 10 15

Pro Cys Tyr Leu Lys Ser Met Tyr Gln Ser Arg Gly Ile Tyr Leu Asn
20 25 30

Ala Lys

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 38 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 1..38
 - (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1150-1263 of SEQ ID NO 2."
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Val Ala Phe Cys Ile His Asn Ile Ala Tyr Gln Gly Arg Phe Ser Phe
1 5 10 15

Ser Asp Phe Pro Leu Leu Asn Leu Pro Asp Glu Phe Arg Gly Ser Phe
20 25 30

Asp Phe Ile Asp Gly Tyr
35

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 79 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (ix) FEATURE:
- (A) NAME/KEY: Modified-site
 - (B) LOCATION: 1..79
 - (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1349-1585 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Lys Pro Val Lys Gly Arg Lys Ile Asn Trp Met Lys Ala Gly Ile Leu
1 5 10 15

Glu Ser His Arg Val Val Thr Val Ser Pro Tyr Tyr Ala Gln Glu Leu
20 25 30

Val Ser Ala Val Asp Lys Gly Val Glu Leu Asp Ser Val Leu Arg Lys
35 40 45

Thr Cys Ile Thr Gly Ile Val Asn Gly Met Asp Thr Gln Glu Trp Asn
50 55 60

Pro Ala Thr Asp Lys Tyr Thr Asp Val Lys Tyr Asp Ile Thr Thr
65 70 75

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 1..59
(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1676-1855 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Val Met Asp Ala Lys Pro Leu Leu Lys Glu Ala Leu Gln Ala Ala Val
1 5 10 15

Gly Leu Pro Val Asp Lys Lys Ile Pro Leu Ile Gly Phe Ile Gly Arg
20 25 30

Leu Glu Glu Gln Lys Gly Ser Asp Ile Leu Ala Val Ala Ile His Lys
35 40 45

Phe Ile Gly Leu Asp Val Gln Ile Val Val Leu
50 55

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 1..64
(D) OTHER INFORMATION: /note= "Amin acid sequence encoded by nucleotides 1945-2136 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Gly Thr Gly Lys Lys Glu Phe Glu Gln Glu Ile Glu Gln Leu Glu Val
1 5 10 15
Leu Tyr Pro Asn Lys Ala Lys Gly Val Ala Lys Phe Asn Val Pro Leu
20 25 30
Ala His Met Ile Thr Ala Gly Ala Asp Phe Met Leu Val Pro Ser Arg
35 40 45
Phe Glu Pro Cys Gly Leu Ile Gln Leu His Ala Met Arg Tyr Gly Thr
50 55 60

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..29
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 2301-2386 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Val Pro Ile Cys Ala Ser Thr Gly Gly Leu Val Asp Thr Val Lys Glu
1 5 10 15
Gly Tyr Thr Gly Phe His Met Gly Ala Phe Asn Val Glu
20 25

B1
W
(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..19
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 2492-2459 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Cys Asp Val Val Asp Pro Ala Asp Val Leu Lys Ile Val Thr Thr Val
1 5 10 15
Ala Arg Ala

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 111 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..111
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1200-1532 of SEQ ID NO 5."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Ala Ser Ile Thr Ala Ser His His Phe Val Ser Arg Ser Gln Thr
1 5 10 15

Ser Leu Asp Thr Lys Ser Thr Leu Ser Gln Ile Gly Leu Arg Asn His
20 25 30

Thr Leu Thr His Asn Gly Leu Arg Ala Val Asn Lys Leu Asp Gly Leu
35 40 45

Gln Ser Thr Thr Asn Thr Lys Val Thr Pro Lys Met Ala Ser Arg Thr
50 55 60

Glu Thr Lys Arg Pro Gly Cys Ser Ala Thr Ile Val Cys Gly Lys Gly
65 70 75 80

Met Asn Leu Ile Phe Val Gly Thr Glu Val Gly Pro Trp Ser Lys Thr
85 90 95

Gly Gly Leu Gly Asp Val Leu Gly Gly Leu Pro Pro Ala Leu Ala
100 105 110

(2) INFORMATION FOR SEQ ID NO:19:

B1
cont
(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..43
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 3817-3945 of SEQ ID NO. 5."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Cys Asp Val Val Asp Pro Ala Asp Val Leu Lys Ile Val Thr Thr Val
1 5 10 15

Ala Arg Ala Leu Ala Val Tyr Gly Thr Leu Ala Phe Ala Glu Met Ile
20 25 30

Lys Asn Cys Met Ser Glu Glu Leu Ser Trp Lys
35 40

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: amin acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..38
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 4031-4144 of SEQ ID NO. 5."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Glu Pro Ala Lys Lys Trp Glu Thr Leu Leu Leu Gly Leu Gly Ala Ser
1 5 10 15

Gly Ser Glu Pro Gly Val Glu Gly Glu Ile Ala Pro Leu Ala Lys
20 25 30

Glu Asn Val Ala Thr Pro
35

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(ix) FEATURE:

- (A) NAME/KEY: misc_RNA
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /note= "Nucleotide 1 is a 7-methyl guanine added by 5'-5' linkage as an RNA cap."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GAUGGCAAGA AAAAAAA

17